

**EXAMINER'S AMENDMENT**

An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no later than the payment of the issue fee.

Robert Kang (Reg. No. 60,135), the attorney in record, gave authorization for this Examiner's Amendment over the telephone during an interview. The claims amendments are as follow:

**PLEASE AMEND THE CLAIMS AS FOLLOWS:**

1. (Currently amended) A method of serving data to a plurality of clients in a client-server environment, comprising the steps of:

generating a plurality of versions of given data in which at least two versions of the given data have different overheads associated therewith, the overhead of a given version of the given data comprising a quantity of processing resources required to serve the given version of the given data;

assigning individual clients to one of a plurality of quality-of-service classes; and satisfying requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve while a client belonging to a low quality-of-service class receives a data version which requires lower overhead to serve;

wherein the overhead to serve a version is correlated with a quality of the version;

wherein the plurality of versions comprise images of different resolutions and clients belonging to the high quality-of-service class are given preferential access to higher resolution images while a client belonging to the low quality-of-service class receives a lower resolution image;

wherein the quality of a version is correlated with a processing time required to create the version; and

wherein the steps are performed at least in part by at least one processor.

2-4. (Canceled)

5. (Original) The method of claim 1, wherein the overhead to serve a version is correlated with how current the version is.

6. (Original) The method of claim 1, further comprising the step of:  
in response to a system load exceeding a threshold, satisfying a higher percentage of requests from clients belonging to a lower quality-of-service class with a version requiring lower overhead to serve.

7. (Original) The method of claim 1, wherein the server comprises multiple nodes and different nodes provide data versions requiring different overheads to serve.

8. (Original) The method of claim 1, further comprising the step of implementing a quality-of-service policy that specifies at least one of content quality and latency.

9. (Original) The method of claim 8, wherein one or more clients belonging to a premium service class are served with high content quality and low latency.

10. (Original) The method of claim 8, wherein one or more clients belonging to a medium service class are served with one of high content quality and low latency.

11. (Original) The method of claim 8, wherein one or more clients belonging to a best-effort service class are served with unspecified content quality and latency.

12. (Original) The method of claim 1, wherein a client request is routed using at least one of an identity of the client, a quality of content, a load on at least one server, a data distribution on at least one server, and a capacity of at least one server.

13. (Original) The method of claim 1, wherein a client is assigned to a quality-of-service class by program logic that is externalized from the server.

14. (Canceled)

15. (Original) The method of claim 1, further comprising the step of satisfying requests using a policy determined by program logic that is externalized from the server.

16. (Canceled)

17. (Currently amended) Apparatus for serving data to a plurality of clients in a client-server environment, comprising:

a memory, and

at least one processor coupled to the memory and operative to: (i) generate a plurality of versions of given data in which at least two versions of the given data have different overheads associated therewith, the overhead of a given version of the given data comprising a quantity of processing resources required to serve the given version of the given data; (ii) assign individual clients to one of a plurality of quality-of-service classes; and (iii) satisfy requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve while a client belonging to a low quality-of-service class receives a data version which requires lower overhead to serve;

wherein the overhead to serve a version is correlated with a quality of the version;

wherein the plurality of versions comprise images of different resolutions and clients belonging to the high quality-of-service class are given preferential access to higher resolution images while a client belonging to the low quality-of-service class receives a lower resolution image; and

wherein the quality of a version is correlated with a processing time required to create the version.

18-20. (Canceled)

21. (Original) The apparatus of claim 17, wherein the overhead to serve a version is correlated with how current the version is.

22. (Original) The apparatus of claim 17, wherein the at least one processor is further operative to, in response to a system load exceeding a threshold, satisfy a higher percentage of requests from clients belonging to a lower quality-of-service class with a version requiring lower overhead to serve.

23. (Original) The apparatus of claim 17, wherein the at least one processor comprises multiple nodes and different nodes provide data versions requiring different overheads to serve.

24. (Original) The apparatus of claim 17, wherein the at least one processor is further operative to implement a quality-of-service policy that specifies at least one of content quality and latency.

25. (Original) The apparatus of claim 24, wherein one or more clients belonging to a premium service class are served with high content quality and low latency.

26. (Original) The apparatus of claim 24, wherein one or more clients belonging to a medium service class are served with one of high content quality and low latency.

27. (Original) The apparatus of claim 24, wherein one or more clients belonging to a best-effort service class are served with unspecified content quality and latency.

28. (Original) The apparatus of claim 17, wherein a client request is routed using at least one of an identity of the client, a quality of content, a load on at least one server, a data distribution on at least one server, and a capacity of at least one server.

29. (Currently amended) An article of manufacture for use in serving data to a plurality of clients in a client-server environment, comprising a non-transitory machine readable storage medium containing one or more programs which when executed implement the steps of:

generating a plurality of versions of given data in which at least two versions of the given data have different overheads associated therewith, the overhead of a given version of the given data comprising a quantity of processing resources required to serve the given version of the given data;

assigning individual clients to one of a plurality of quality-of-service classes; and  
satisfying requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve while a client belonging to a low quality-of-service class receives a data version which requires lower overhead to serve;

wherein the overhead to serve a version is correlated with a quality of the version;  
wherein the plurality of versions comprise images of different resolutions and clients belonging to the high quality-of-service class are given preferential access to higher resolution images while a client belonging to the low quality-of-service class receives a lower resolution image; and

wherein the quality of a version is correlated with a processing time required to create the version.

30. (Currently amended) A system, comprising:  
a plurality of clients, each client belonging to a quality-of-service class;  
a load balancer for sending requests from clients to at least one back-end server; and  
at least one back-end server for (i) generating a plurality of versions of a given object[[s]] in which at least two versions of [[a]] the given object have different overheads associated therewith, the overhead of a given version of the given object comprising a quantity of processing resources required to serve the given version of the given object and (ii) satisfying the requests so that a client belonging to a high quality-of-service class is given preferential access to object versions which require higher overheads to serve while a client belonging to a low quality-of-service class receives an object version which requires lower overhead to serve;

wherein the overhead to serve a version is correlated with a quality of the version;

wherein the plurality of versions comprise images of different resolutions and clients belonging to the high quality-of-service class are given preferential access to higher resolution images while a client belonging to the low quality-of-service class receives a lower resolution image;

wherein the quality of a version is correlated with a processing time required to create the version; and

wherein at least a portion of the system comprises a memory and at least one processor coupled to the memory.

31. (Currently amended) A method of providing a data serving service, comprising the step of:

a service provider: (i) generating a plurality of versions of given data in which at least two versions of the given data have different overheads associated therewith, the overhead of a given version of the given data comprising a quantity of processing resources required to serve the given version of the given data; (ii) assigning individual clients to one of a plurality of quality-of-service classes; and (iii) satisfying requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve while a client belonging to a low quality-of-service class receives a data version which requires lower overhead to serve;

wherein the overhead to serve a version is correlated with a quality of the version;

wherein the plurality of versions comprise images of different resolutions and clients belonging to the high quality-of-service class are given preferential access to higher resolution images while a client belonging to the low quality-of-service class receives a lower resolution image; wherein the quality of a version is correlated with a processing time required to create the version; and

wherein the step is performed at least in part by at least one processor.

32. (Original) The method of claim 31, wherein the data serving service comprises a quality-of-service policy specification.

33. (Original) The method of claim 32, wherein the quality-of-service policy specification comprises:

a plurality of subscriptions, each subscription being specified by content quality and service latency, wherein a limited premium service subscription is served with high content quality in low service latency, a medium service subscription is served with a high content quality or a low service latency, and an unlimited best-effort service subscription is served with unspecified content quality and latency.

34. (Original) The method of claim 31, wherein the service provider modifies data content and how the data content is served to clients in response to one or more changing conditions.

35. (Original) The method of claim 34, wherein one or more changing conditions comprises a source of a bottleneck.

36. (Original) The method of claim 31, wherein the step of assigning individual clients to one of a plurality of quality-of-service classes is based on a client payment.

37. (Currently amended) A method of serving data to a plurality of clients, comprising the steps of:

generating a plurality of versions of given data in which at least two versions of the given data have different overheads associated therewith, the overhead of a given version of the given data comprising a quantity of processing resources required to serve the given version of the given data;

establishing at least two quality-of-service classes; and

satisfying requests so that a client belonging to one quality-of-service class is served with one version of the given data having one overhead associated therewith, while a client belonging to another quality-of-service class is served with another version of the given data having another

overhead associated therewith, ~~the overhead of a given version of the given data comprising a quantity of processing resources required to serve the given version of the given data;~~

wherein the overhead of the given version of the given data is correlated with a quality of the given version;

wherein the plurality of versions of the given data comprise images of different resolutions and clients belonging to the high quality-of-service class are given preferential access to higher resolution images while a client belonging to the low quality-of-service class receives a lower resolution image;

wherein the quality of the given version is correlated with a processing time required to create the given version; and

wherein the steps are performed at least in part by at least one processor.

38. (Previously presented) The system of claim 30, where the at least one back-end server comprises:

at least a first back-end server for generating a first version of the given object; and

at least a second back-end server for generating a second version of the given object;

wherein the first and second versions of the given object have different overheads associated therewith.

39. (Previously presented) The method of claim 37, wherein the one version of the given data is served by one back-end server while the other version of the given data is served by another back-end server.

### ***Allowable Subject Matter***

Claims 1, 5-13, 15, 17 and 21-39 are allowed.

The following is an Examiner's statement of reasons for allowance:

The prior arts in record fail to teach "generating a plurality of versions of given data in which at least two versions of the given data have different overheads



associated therewith, the overhead of a given version of the given data comprising a quantity of processing resources required to serve the given version of the given data, assigning individual clients to one of a plurality of quality-of-service classes, and satisfying requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve while a client belonging to a low quality-of-service class receives a data version which requires lower overhead to serve, wherein the overhead to serve a version is correlated with a quality of the version, wherein the plurality of versions comprise images of different resolutions and clients belonging to the high quality-of-service class are given preferential access to higher resolution images while a client belonging to the low quality-of-service class receives a lower resolution image, wherein the quality of a version is correlated with a processing time required to create the version, and wherein the steps are performed at least in part by at least one processor," as recited in independent claim 1.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

### ***Conclusion***

Any inquiry concerning this communication or earlier communications from the examiner should be directed to El Hadji M Sall whose telephone number is 571-272-4010. The examiner can normally be reached on 8:00-4:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Ario Etienne can be reached on 571-272-4001. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

/El Hadji M Sall/

Examiner, Art Unit 2457

/ARIO ETIENNE/

Supervisory Patent Examiner, Art Unit 2457